PATENT APPLICATION

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE HONORABLE BOARD OF PATENT APPEALS AND INTERFERENCES

In re the Application of

James M. Sweet et al.

Confirmation No. 8445

Application No.: 10/608,591

Examiner: Nathan Hillery

Filed: 06/27/2003

Docket No.: A2555Q1-US-NP

For: **DETERMINATION OF TABLE OF CONTENT LINKS FOR A HYPERLINKED DOCUMENT**

BRIEF ON APPEAL

Appeal from Group 2176

Christopher D. Wait
XEROX CORPORATION
Xerox Square - 20A
Rochester, NY 14644
Telephone: (585) 423-6918
Attorneys for Appellants

TABLE OF CONTENTS

I. REAL PARTY IN INTEREST

The real party in interest for this appeal and the present application is Xerox Corporation, by way of an Assignment recorded in the U.S. Patent and Trademark Office at Reel 14257, Frame 925-927 and Reel 014557, Frame 0676-0677.

II. STATEMENT OF RELATED APPEALS AND INTERFERENCES

Following are identified any prior or pending appeals, interferences or judicial proceedings, known to Appellant, Appellant's representative, or the Assignee, that may be related to, or which will directly affect or be directly affected by or have a bearing upon the Board's decision in the pending appeal:

Appeal Brief filed in copending Application No. 10/608,590

Appeal Brief to be filed in copending Application No. 10/608,587

There are no further prior or pending appeals, interferences or judicial proceedings, known to Appellant, Appellant's representative, or the Assignee, that may be related to, or which will directly affect or be directly affected by or have a bearing upon the Board's decision in the pending appeal.

III. STATUS OF CLAIMS

Claims 1, 2, 4, 6, 7, 9, 11, 12, and 14 are on appeal.

Claims 1-15 are rejected.

IV. STATUS OF AMENDMENTS

No Amendment After Final Rejection has been filed.

V. SUMMARY OF CLAIMED SUBJECT MATTER

The subject matter of independent claims 1, 6, and 11, is directed to a methodology for assembling a document from content spanning multiple web-pages by employing two cooperative processes. Given a starting location 110, one process analyzes a single page at a time to find candidate links 140. The links are recursively followed and those pages are analyzed. A detailed set of heuristics is used to determine what is or is not a candidate link. The links are examined for link clusters and a table of contents if found is identified. The candidate pages 120 are then fed to a document-level analyzer 150. This process compares the attributes of one page against the others and looks for a document-like structure. Using another detailed set of heuristics, the document-level analyzer 150 determines if the page should be included in the document. (see Abstract, page 20 of the specification as filed, and Figure 1) [in support of claims 1, 6, and 11]

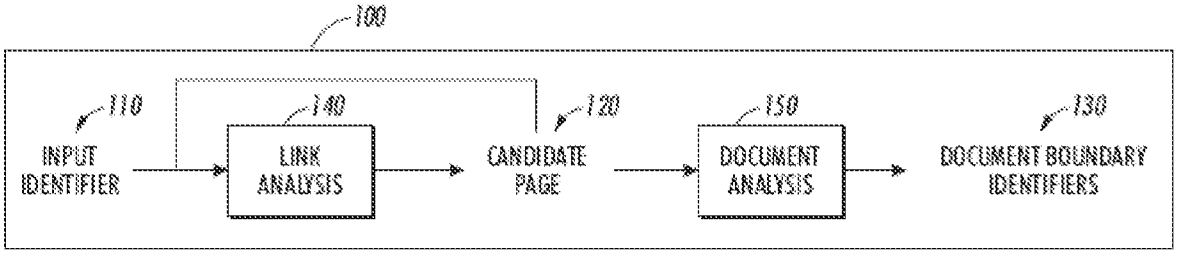

FIG. 1

In more particular support of the subject matter of independent claims 6, and 11, the page-level link analysis 140 is described in greater detail in Figure 2. During page-level link analysis 140, the document detection system attempts to identify links that may potentially lead to other pages within the same document. It is assumed that a well-authored multi-page document will always include progression links (links that

provide some well-defined progression through the document, often indicated by the presence of some well-known contextual clue, such as a graphic or text “next” or “previous” indicator) and/or table of contents links (clusters of links providing a path to every page or some logical subset of pages in the document) that indicate the structure of the document. These are the two categories of intra-document links that the link analysis process 140 seeks to identify. (see page 7, lines 10-20 of the specification as filed, and Figure 2) [in support of claims 6, and 11]

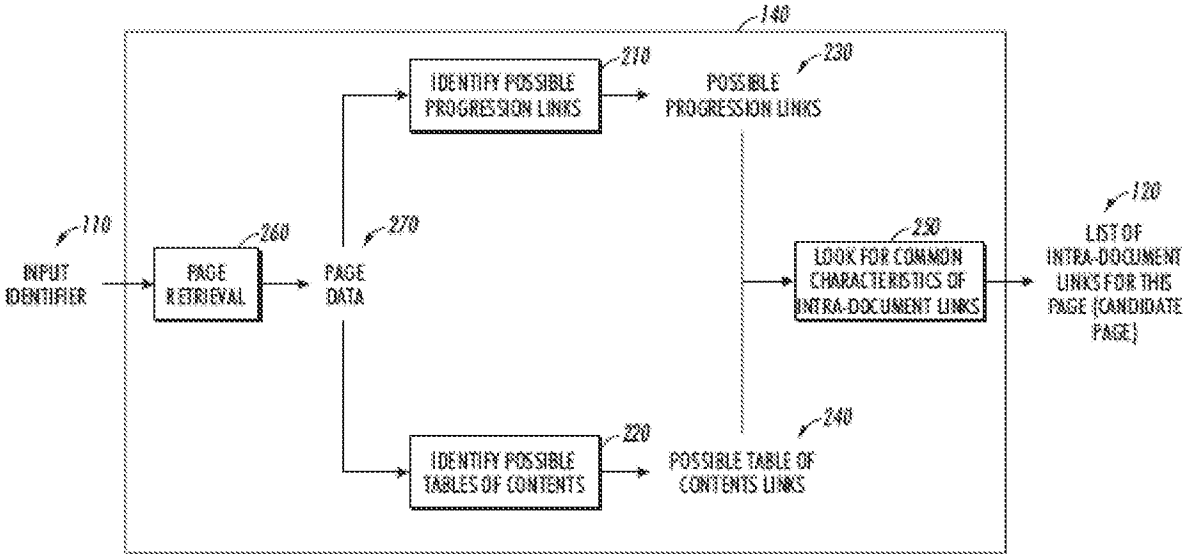

FIG. 2

In further support of the subject matter of independent claims 6, and 11, the link analysis process begins with the retrieval of the actual page 270 for analysis from the page identifier 110. This is done as will be well understood by those skilled in the art, by the page retrieval process 260. The retrieved page 270 is then used as input to both the progression-link identification module 210 and the link-cluster identification module 220. In the progression-link identification module 210, possible progression links 230 are identified primarily by means of a progression indicator, which is a textual or

graphical clue that suggests the nature of the progression link. Link-cluster identification module 220 examines the page data 270 to identify link clusters and thereby possible table of content type links 240. The possible progression links 230 and possible table of content links 240 are passed to module 250 for a final examination to weed out links which have properties that are not characteristic of typical intra-document links, e.g. they point to a different web server. The final result is then a list of intra-document links 120 for the candidate page 270. (see page 7, lines 22-35 of the specification as filed, and Figure 2)

Figure 2, module 220 examines the page data 270 to identify link clusters. It is assumed that in a well-authored hypertext page, table of contents links will appear in clusters, thereby indicating to the user that all of these links are part of a single cohesive construct. Given this assumption, the first step in locating a table of contents is to locate all of the link clusters in a particular page.

The Identification of link clusters is based on three criteria:

1) Proximity: The links in a cluster should be close together. The same heuristic as applied to identification of the most proximal link for a progression indicator can be used here to identify groups of links that have a low perceived distance.

2) Similarity: The links in a cluster should look like each other, i.e. they will usually all be of the same font, type size, and color.

3) Regularity: If there is intervening content between the links, or if the links are dissimilar, these lapses in Proximity and Similarity should form some sort of consistent pattern. One example is a table of contents where each link has a chapter description below it (Proximity is low, but the pattern of intervening content is highly consistent). Another example is a table of links where the color of the text alternates in

each column in order to make it more readable (Similarity is low, but the changes in appearance form a simple pattern).

Regularity is measured by performing pattern matching on the intervening content and document structure tags between pairs of nearby links. The other two criteria are easily measured by simple heuristics.

Once all link clusters in a web page have been identified, the task remains of distinguishing which clusters represent tables of contents and which represent other constructs, such as navigation bars or bibliographies. The primary determining criteria for this is the similarity between the link targets of the links in the cluster, i.e. collocation on the same server, residence in the same directory or nearby area of the directory hierarchy, and similarity in filename. (see page 10, lines 4-33 of the specification as filed) [in support of claims 6, and 11]

VI. GROUNDS OF REJECTION TO BE REVIEWED ON APPEAL

The following grounds of rejection are presented for review:

Claims 1, 2, 4, 6, 7, 9, 11, 12, and 14 are rejected under 35 U.S.C. §103(a) as being unpatentable over U.S. Patent No. 6,112,203 to Bharat et al. (hereinafter Bharat) in further view of U.S. Patent No. 5,924,104 to Earl (hereinafter Earl).

VII. ARGUMENT

A. Claims 1, 2, 4, 6, 7, 9, 11, 12, and 14 Would Not Have Been Obvious Over Bharat in View of Earl

Claims 1, 2, 4, 6, 7, 9, 11, 12, and 14 are rejected under 35 U.S.C. §103(a) as being unpatentable over Bharat in further view of Earl.

Problematically, neither Bharat or Earl, alone or in combination, teach or suggest the Applicants' invention. Claim elements are missing from the Bharat and Earl references. Indeed the references teach away from the Applicants' claimed invention. A Prima facie case for Obviousness has thus not been made out. Further, no finding has been provided directed to: a identifiable reason that would have prompted a person of ordinary skill in the relevant field to combine the elements in the way that the Applicants' claimed new invention does. Thus, the Applicant is faced with the conundrum of positively proving a negative. That is in other words: proving that something which is not there, is not there.

Bharat teaches that in a computerized method, a set of documents is ranked according to their content and their connectivity by using topic distillation. The documents include links that connect the documents to each other, either directly, or indirectly. A graph is constructed in a memory of a computer system. In the graph, nodes represent the documents, and directed edges represent the links. Based on the number of links connecting the various nodes, a subset of documents is selected to form a topic. A second subset of the documents is chosen based on the number of directed edges connecting the nodes. Nodes in

the second subset are compared with the topic to determine similarity to the topic, and a relevance weight is correspondingly assigned to each node. Nodes in the second subset having a relevance weight less than a predetermined threshold are pruned from the graph. The documents represented by the remaining nodes in the graph are ranked by connectivity based ranking scheme.

It is essential to the understanding Bharat that *Bharat is directed to a search engine* and as such is sorting through pages already identified by a simple word string search (please see column 1, lines 14-54 of Bharat). Bharat is concerned with solving the problem of answering a search engine query, and thus with ranking a set of documents to point to in response to that query. The Applicants however, are teaching that having identified where one document page is, how to find and pull together all relevant pages associated with that document into a single coherent document (please see page 5, first paragraph, of the Applicants' specification). That is, a single coherent document representation suitable for printing and downloaded viewing. As such the Applicants teach "to weed out links which have properties that are not characteristic of ***intra***-document links" and thus eschew all other documents. Bharat on the other hand, will not link (i.e. Bharat will reject) self referencing pages so as not to unduly influence the search outcome (see column 5, lines 17-20) where Bharat provides:

> "If a link points to a page that is represented by a node in the graph, and both pages are on different servers, then a corresponding edge 213 is added to the graph 211. ***Nodes representing pages on the same server are not linked. This prevents a single Web site with many self-***

referencing pages to unduly influence the outcome. This completes the n-graph 211."

Thus Bharat is interested in only ***inter***-document links for the sake of ranking links. Bharat does not assemble a single coherent document but a link list of search results responsive to a word query. Thus Bharat does NOT examine "the collective set of identified candidate document pages to weed out links which have properties that are not characteristic of ***intra***-document links". Thus a claim element is missing.

Indeed, Bharat teaches away from the Applicants' invention. The Applicants teach to embrace that which Bharat discards. The cited text from the Applicants' specification page 7, lines 22-35 follows:

"The link analysis process begins with the retrieval of the actual page 270 for analysis from the page identifier 110. This is done as will be well understood by those skilled in the art, by the page retrieval process 260. The retrieved page 270 is then used as input to both the progression-link identification module 210 and the link-cluster identification module 220. In the progression-link identification module 210, possible progression links 230 are identified primarily by means of a progression indicator, which is a textual or graphical clue that suggests the nature of the progression link. Link-cluster identification module 220 examines the page data 270 to identify link clusters and thereby possible table of content type links 240. The possible progression links 230 and possible table of content links 240 are passed to module 250 for a final examination to weed out links which have properties that are not characteristic of

typical intra-document links, e.g. they point to a different web server. The final result is then a list of intra-document links 120 for the candidate page 270."

To paraphrase, the possible links are passed by the Applicants, to weed out those links which are not characteristic of typical intra-document links. An example of the links which are weeded out are those which point to a different web server. (please see also page 10, lines 32-34, of the Application Specification as filed). These links are not likely to be part of the document the Applicants are trying to assemble. Bharat does just the opposite, as noted above, so as to not unduly influence the outcome of the results to the user query. Please also see the attached §132 Declarations, particularly in the Sweet Declaration, item numbers 7-11, and in the Harrington Declaration, item number 7.

Earl fails to provide what Bharat lacks, nor does it provide any teaching relating to the Applicants' claimed invention. Earl provides link lists like Bharat but provides different presentation styles for the links to a user depending on whether they are intra-document or inter-document. Actually what Earl defines as intra-document is what the Applicants would call intra-page, i.e. a link pointing to a location somewhere further down the same page. And thus what Earl calls inter-document is really inter-page. The teaching found in Earl is simply about providing some indicia to the viewer as to whether a hyper link will take the user elsewhere down the present page or to an entirely different page.

Please see the attached §132 Declarations for what one skilled in the art would consider to be "intra-document" versus intra-page, particularly in the Sweet

Declaration, item numbers 13-15, and in the Harrington Declaration, item number 5. The Applicants are teaching assembling a document, and having identified the current page, have no interest in self referential links to that same page (they already have it) and thus would discard, or weed out, those links which Earl keeps.

Earl, having made a discrimination between two type of links, keeps all those links, choosing only to display them differently. The Applicants having discriminated between links to find some as not pointing to more of the desired document, discard or weed out or filter out those links. A gardener, weeding out a flower bed and having spotted a weed, does not keep that weed in their flower bed to display differently. But Earl does. Thus Earl does NOT examine "the collective set of identified candidate document pages to weed out links which have properties that are not characteristic of typical intra-document links".

In rebuttal to this previously presented argument above analogizing the terminology "weed out" to gardening, the Examiner has asserted that the Office "is forced to rely on the knowledge of one of ordinary skill in the art". The Applicants must emphatically traverse as to how one skilled in the art would interpret this terminology. Please see the attached §132 Declarations particularly the Sweet Declaration, item number 8 and especially item 9 for what those skilled in the art would consider to be meant by the terminology "weed out". Earl only discriminates, but does not discard, thus Earl does not "weed out".

It must also be pointed out that neither Bharat or Earl concern themselves with the claim element of a table of contents. This is again not a surprising finding as they are both directed to search inquiries rather than the singular

document image which the Applicants endeavor to assemble. Finding a table of contents is an important find in tracing out a single hyperlinked document, but of little consequence to a search engine inquiry. Thus yet another of the Applicants' claim elements is absent in the cited art of Bharat and Earl.

Therefore, Earl in turn fails to provide the elements that Bharat also lacks, and the combination of Bharat and Earl thus fails to provide the requirements for a Prima Facie case of obviousness and the rejection is improper.

Further, no finding has been provided by the Examiner directed to some *genuine* identifiable reason that would have prompted a person of ordinary skill in the relevant field to combine the elements in the way that the Applicants' claimed new invention does. The importance of doing so is clearly stated in *KSR*, 550 U.S., 82 USPQ2d at 1395 and 1396. It would appear that the Examiner is using Appellant's disclosure as a recipe for selecting the appropriate portions of the prior art to construct Appellant's claimed invention. A piecemeal reconstruction of prior art patents in light of Appellant's disclosure should not be a basis for a holding of obviousness, especially when claim elements are absent.

VIII. CONCLUSION

For all of the reasons discussed above, it is respectfully submitted that the rejections are in error and that claims 1, 2, 4, 6, 7, 9, 11, 12, and 14 are in condition for allowance. For all of the above reasons, Appellants respectfully request this Honorable Board to reverse the rejections of claims 1, 2, 4, 6, 7, 9, 11, 12, and 14.

Respectfully submitted,

/Christopher D. Wait, Reg. #43230/
Christopher D. Wait
Attorney for Applicant(s)
Registration No. 43,230
Telephone (585) 423-6918

XEROX CORPORATION
Xerox Square – 20A
Rochester, NY 14644

Filed: October 10, 2008

CLAIMS APPENDIX

CLAIMS INVOLVED IN THE APPEAL:

1. (Previously Presented) An automated identification methodology for identification of table of content links in a given hyperdocument for assembling a document representation by gathering the content of hyperlinked pages pointed to by the identified table of contents comprising:

searching page data to create a list of links in the given hyperdocument;

analyzing each link in conjunction with each other link in the list of links to identify link pairings;

assembling link pairings in order to form clusters of links;

examining the links in the cluster of links for locality;

weeding out the links from the cluster of links which have properties that are not characteristic of intra-document links, to provide a resultant table of content set of identified candidate document pages; and,

grouping the content found in the resultant table of content set of candidate document pages by an automated system into a document representation stored in memory by the automated system; and,

printing, or viewing on a display by a user, the document representation.

2. (Original) The method of claim 1 wherein the step for analyzing each link further comprises determining a score for each link pairing.

3. (Original) The method of claim 2 wherein the scoring is determined by a proximity criteria.

4. (Original) The method of claim 2 wherein the scoring is determined by a similarity criteria.

5. (Original) The method of claim 2 wherein the scoring is determined by a regularity criteria.

6. (Previously Presented) A system identification methodology for assembling a document representation for subsequent viewing or printing of a given hyperlinked hyperdocument by gathering related hyperlinked page content comprising:

performing a page-level link analysis that identifies those hyperlinks on a page linking to a candidate document page further comprising a methodology of:

analyzing each link in conjunction with each other link to identify link pairings;

assembling link pairings in order to form clusters of links; and,

examining the links in the cluster of links for locality;

performing a recursive application of the page-level link analysis to the linked candidate document page and any further nested candidate document pages thereby identified, until a collective table of content set of identified candidate document pages is assembled;

performing a document-level analysis that examines the collective table of content set of identified candidate document pages for grouping into one or more documents;

examining the collective table of content set of identified candidate document pages to weed out links from the collective table of content set which have properties that are not characteristic of intra-document links, to provide a resultant set of identified candidate document pages; and,

grouping the content found in the resultant set of candidate document pages by an automated system into a document representation stored in memory by the automated system; and,

printing, or viewing on a display by a user, the document representation.

7. (Original) The method of claim 6 wherein the step for analyzing each link further comprises determining a score for each link pairing.

8. (Original) The method of claim 7 wherein the scoring is determined by a proximity criteria.

9. (Original) The method of claim 7 wherein the scoring is determined by a similarity criteria.

10. (Original) The method of claim 7 wherein the scoring is determined by a regularity criteria.

11. (Previously Presented) A system identification methodology for assembling a document representation for subsequent viewing or printing of a given hyperlinked hyperdocument by gathering related hyperlinked page content comprising:

performing a page-level link analysis that identifies those hyperlinks on a page linking to a candidate document page further comprising a methodology of:

searching page data to create a list of links in the hyperdocument;

analyzing each link in conjunction with each other link in the list of links to identify link pairings;

assembling link pairings in order to form clusters of links; and,

examining the links in the cluster of links for locality;

performing a recursive application of the page-level link analysis to the linked candidate document page and any further nested candidate document pages thereby identified, until a collective table of content set of identified candidate document pages is assembled; and,

performing a document-level analysis that examines the collective table of content set of identified candidate document pages for grouping into one or more documents

examining the collective table of content set of identified candidate document pages to weed out links from the collective table of content set which have properties that are not characteristic of intra-document links, to provide a resultant set of identified candidate document pages; and,

grouping the content found in the resultant set of candidate document pages by an automated system into a document representation stored in memory by the automated system; and,

printing, or viewing on a display by a user, the document representation.

12. (Original) The method of claim 11 wherein the step for analyzing each link further comprises determining a score for each link pairing.

13. (Original) The method of claim 12 wherein the scoring is determined by a proximity criteria.

14. (Original) The method of claim 12 wherein the scoring is determined by a similarity criteria.

15. (Original) The method of claim 12 wherein the scoring is determined by a regularity criteria.

EVIDENCE APPENDIX

A copy of each of the following items of evidence relied on by the Appellant is attached:

DECLARATION UNDER 37 CFR §1.132 by Steven J. Harrington, Ph. D., filed 11/16/07.
The evidence was entered into the record by the Examiner in the 01/09/08 Office Action.

DECLARATION UNDER 37 CFR §1.132 by James M. Sweet, filed 11/16/07.
The evidence was entered into the record by the Examiner in the 01/09/08 Office Action.

RELATED PROCEEDINGS APPENDIX

NONE